

Minimum Error Discrimination of Linearly Independent Pure States: Analytic Properties of POVM

Tanmay Singal^{*1} and Sibasish Ghosh^{†1}

¹Optics and Quantum Information Group, Institute of Mathematical Sciences, CIT Campus, Taramani, Chennai, 600 113, India

February 20, 2014

Abstract

The optimization conditions for minimum error discrimination of linearly independent pure states comprise of two kinds: stationary conditions over the space of rank one projective measurements and the global maximization conditions. A discrete number of projective measurements will solve the former of which a unique one will solve the latter. In the case of three real linearly independent pure states we show that the stationary conditions translate to a system of simultaneous polynomial (non linear) equations in three variables thus explaining why it's so difficult to obtain a closed-form solution for the optimal POVM. Additionally, our method suggests that as an ensemble of LI pure states is varied as a smooth function of some independent parameters, the optimal POVM will also vary smoothly as a function of the same parameters. By employing the implicit functions theorem we exploit this fact to obtain a technique to find the solution of MED of LI pure states by dragging the solution from a known example (say, pure orthogonal states) to any general linearly independent ensemble of pure states in the same Hilbert space. By employing RK4 to solve the first order coupled non-linear differential equations find that the resulting error is within the RK4 error performance.

1 Introduction

Minimum Error Discrimination is one of the oldest problems in quantum state discrimination. The problem arises due to the vectorial nature of states which makes them indistinguishable through measurement. To infer what the state is, one has to perform measurement. Non-orthogonality of states implies that the measurement operators cannot perfectly distinguish one state from the other leading to errors in the discrimination procedure. Different measurement strategies have different performance strength (measured in terms of the average probability of error or the average probability of success). Given that the states cannot be distinguished perfectly there must be some measurement criterion which one could adopt to attain a maximum probability of success. To find what this measurement strategy is, is the problem of quantum state discrimination.

The setting in minimum error discrimination or quantum hypothesis testing or ambiguous state discrimination is the following: A has a fixed ensemble of states $\{\rho_1, \rho_2, \dots, \rho_m\}$ from which she selects one with certain probability $\{p_1, p_2, \dots, p_m\}$ ($p_i > 0$, $\sum_i p_i = 1$) respectively and gives it to B. B knows that A has selected this state from that fixed set with those probabilities and his job now figure out which state he has been given using an m element POVM. There is a one-to-one correspondence between elements in A's ensemble $\{\rho_i, p_i\}_{i=1}^m$ and B's POVM elements $\{\Pi_i\}_{i=1}^m$ so that when the i th measurement outcome clicks, B infers A gave him the i th state from her ensemble. One can infer that

^{*}stanmay@imsc.res.in

[†]sibasish@imsc.res.in

the non-orthogonality of the ensemble elements implies that errors are likely to occur. B's job is to now find the optimal POVM for maximizing the average probability of success of his outcome.

There are other variants to the state discrimination problem[12][19]. The most popular alternative of them is called unambiguous state discrimination. The idea is that the measurement outcomes are now $m+1$ in number where, as in the MED case, there is a one-to-one correspondence between ensemble elements and the first m POVM elements. In this problem, the POVM is designed such that only when A sends B the i th state will the i th POVM element click, otherwise it won't. The trade-off is the $m+1$ th element in the ensemble which is the "inconclusive" outcome result i.e. B can say nothing about the state A sent him when this POVM element clicks. Heuristically one can expect that a set of linearly dependent states cannot be unambiguously discriminated in this manner, an idea that was proven true later[13]. The more recent proposal of maximum confidence measurements is the application of unambiguous state discrimination to linearly dependent ensembles i.e. when linearly dependent then the measurement which maximizes the confidence of the state measured being the correct one.

Coming back to MED, necessary and sufficient conditions which the optimal POVM has (or have) to satisfy were given by Holevo[8] and Yuen[7] independently. The latter cast MED into a linear programming (and now SDP) problem for which numerical solutions can now be approximated within polynomial time. While there are quite a number of numerical techniques to obtain the optimal POVM upto very good approximation[5, 23, 24], for very few ensembles has the MED problem been solved analytically. Compare that with the unambiguous state discrimination for which many more general ensembles have been solved.[12, 15, 16, 17, 14, 18] Some of these include ensemble of two states[1], ensembles whose density matrix is maximally mixed state[7], equiprobable ensembles that lie on the orbit of a unitary[3],[4],[2] and recently three mixed qubit states[9]. For linearly independent pure state ensembles just the two-state ensemble problem has been solved but nobody has been able to give a solution for three linearly independent pure states.

We are motivated to understand what makes the three linearly independent pure state problem so difficult. For this purpose the problem is cast in different settings: (i) First it is cast in a geometric setting i.e. we make use of the geometry of the qutrit bloch sphere [20] to see what the problem looks like. Corresponding to the necessary and sufficient conditions given by Holevo and Yuen we get a set of simultaneous multivariate non-linear polynomial equations in terms of the components of the 8 dimensional qutrit bloch vectors of three states dual to the linearly independent pure states of the ensemble. It can be seen how complicated the problem here when contrasted with the simplicity of the two state discrimination problem which is trivial once the problem is cast in the qubit state space.(ii) Corresponding to the generalized pretty good measurements [21]/ Belavkin weighted square root measurements [22] one can obtain two set of real simultaneous multivariate polynomial equations - the more complicated one of which is essentially an attempt to find the inverse of the mapping $\tilde{p}_i \rightarrow p_i$ given in [21], while the other one is much simpler. In (i) and (ii) each set of real simultaneous multivariate polynomial equations have a discrete solution set of which a subset corresponds to m -element POVMs and of that a proper subset corresponds to the optimal POVM. The other POVMs have a physical interpretation: they are the stationary points of the objective function in the space of extremal m -element POVMs. When $m = 2$, there are only two possible solutions i.e. two such stationary points - one corresponding to the global minima and the other corresponding to the global maxima; and one is obtained from the other by interchanging indices of POVM elements of the latter. As soon as $m=3$, there can be multiple stationary points in the space of extremal POVMs some of which could be local maxima, local minima, saddle points, points of inflection, the global maxima and minimum etc. Thus all this information is encoded in equations. Given this nature of the solution set, it is unlikely that one can find a set of simpler equations to solve the problem. Add to this the problem that the set of reals is not algebraically closed i.e. real polynomial equations can have complex solutions. Thus often some solutions in the solution set are complex which then don't correspond to any POVM.

One of the methods we used to obtain these equations suggests a way to analyze the analytic properties of the optimal POVM as a function of the ensemble i.e. to analyze how ensemble varies when the ensemble of LI pure states is varied. These ensembles are parameterized using their gram matrices. For studying the analytic behaviour, we employ the implicit function theorem which tells us that the optimal POVM will vary analytically as the ensemble is varied according to some parameterization. Using IFT we obtain a set of first order non-linear coupled differential equations using which we can "drag" the optimal POVM from some ensemble to another. We employ RK4 to drag the optimal POVM from an

equiprobable orthogonal ensemble of states to a general linearly independent ensemble. We note that the error increases slower than expected with the number of iterations telling us that this method is reliable to obtain MED for some LI pure ensemble upto desired accuracy.

The paper is divided as follows: First we go into detail about what MED is. Then we give the optimizing conditions and specify what they look like for LI pure ensembles. Then we give a brief description on LI pure ensembles. We formulate the MED problem for linearly independent states using geometry of the qutrit state space and later in what we call the gram matrix method. In both cases we come across simultaneous real multivariate polynomial equations which have a zero dimensional variety but multiple solutions. We note that while some of the solutions are complex and discardable the others correspond to some rank-one projective measurement and infer that all of them are stationary points of the objective function in the space of rank-one projective measurements. Moving on, we find a way to represent linearly independent ensembles using an equivalence class of trace one positive definite matrices and use that along with the implicit function theorem to show that the optimal POVM will vary analytically as a function of the ensemble. Then again using IFT we obtain a set of first order non-linear coupled differential equations which can be used to drag the solution from one ensemble to another. Starting with an equiprobable orthogonal ensemble we employ RK4 to do this "dragging" and we note that we get a decent performance where the errors remain below expected.

2 Minimum Error Discrimination: The Problem

A has a device that prepares a quantum state, ρ_i with probability p_i from a fixed set, $\{\rho_1, \rho_2, \dots, \rho_m\}$. Here $p_i > 0, \forall 1 \leq i \leq m$ and $\sum_{i=1}^m p_i = 1$. Let $\text{supp}(\rho_i) = \mathcal{H}_i$, $\text{rank}(\rho_i) = r_i$. Also, let $\mathcal{H} = \text{span}(\bigcup_{i=1}^m \mathcal{H}_i)$ and $\dim(H) = n$.

A obtains a quantum state ρ_i with probability p_i from this device and gives this state to B without telling the latter which state he's being given. B knows that this state is from A's device and wants to know which state A has given him. For that A has to "distinguish" the state he's been given from the other ones in $\{\rho_i, \rho_j\}_{i,j=1}^m$ by performing a measurement on his state. This measurement is usually a generalized measurement i.e. a POVM. The measurement scheme is as follows: this measurement has m distinct outcomes each of which are indexed in such a way that the i th outcome indicates that he has been given ρ_i .

In case the states $\{\rho_i\}_{i=1}^m$ are pairwise orthogonal i.e. $\text{Tr}(\rho_i \rho_j) = 0, \forall 1 \leq i, j \leq m$, B can choose his POVM to be $\{P_i\}_{i=1}^m$ where P_i is a projector on the subspace \mathcal{H}_i and $P_i P_j = \delta_{i,j} P_i, \forall 1 \leq i, j \leq m$. In such a case $\text{Tr}(P_i \rho_j) = \delta_{i,j}$ leading to *always* correct inference and we say that elements from the ensemble can be distinguished perfectly.

In case the states $\{\rho_i\}_{i=1}^m$ aren't pairwise orthogonal (as is usually the case) there is no measurement that can distinguish perfectly between them. This means that it may so happen that despite being given ρ_i , B's measurement output is j , leading to an error.

The average probability of error is given by:

$$P_e = \sum_{\substack{i,j=1 \\ i \neq j}}^m p_i \text{Tr}(\rho_i \Pi_j) \quad (1)$$

where $\{\Pi_i\}_{i=1}^m$ represents an m element POVM with $\Pi_i \geq 0$ and $\sum_{i=1}^m \Pi_i = \mathbb{1}$.

The average probability of success is given by:

$$P_s = \sum_{i=1}^m p_i \text{Tr}(\rho_i \Pi_i) \quad (2)$$

Both probabilities sum up to 1:

$$P_s + P_e = 1 \quad (3)$$

It is worth noting that P_s (and P_e) is a linear function of the m-element POVM $\{\Pi_i\}_{i=1}^m$ and that the set of m-element POVMs form a compact convex set. From (3) it immediately follows that maximization of P_s over these m-element POVMs implies the minimization of P_e over the same.

$$P_s^{\max} = \text{Max}_{\substack{\{\Pi_i\}_{i=1}^m \\ \Pi_i \geq 0 \\ \sum_i \Pi_i = \mathbb{1}}} P_s = 1 - P_e^{\min} \quad (4)$$

Thus B's task is finding the optimizing m-element POVM. Given that the set of m-element POVMs forms a compact set, we know that such an m-element POVM must exist.

3 The Optimum Conditions

Thus B's task is an optimization problem (4) over a constrained set $\Pi_i \geq 0$, $\sum_i \Pi_i = \mathbb{1}$. To every constrained optimization problem (called the primal problem) there exists a corresponding dual problem which provides a lower bound (if primal problem is a constrained minimization) or an upper bound (if the primal problem is a constrained maximization) to the quantity being optimized in the primal problem. Under certain conditions these bounds are tight implying that one can obtain solution for the primal problem from its dual. We then say that there is no duality gap between both problems.

For MED there is no duality gap and the dual problem can be solved to obtain optimal POVM. This dual problem is given as follows [7]:

$$\text{Min Tr}(Z) \ni Z \geq p_i \rho_i \forall 1 \leq i \leq m \quad (5)$$

Also the optimal m-element POVM will satisfy the complementarity slackness condition:

$$(Z - p_i \rho_i) \Pi_i = \Pi_i (Z - p_i \rho_i) = 0, \forall 1 \leq i \leq m \quad (6)$$

Now summing over i in (6) and using the fact that $\sum_{i=1}^m \Pi_i = \mathbb{1}$ we get:

$$Z = \sum_{i=1}^m p_i \rho_i \Pi_i = \sum_i \Pi_i p_i \rho_i \quad (7)$$

From (6) we get

$$\begin{aligned} \Pi_j (Z - p_i \rho_i) \Pi_i &= \Pi_j (Z - p_j \rho_j) \Pi_i \\ \Rightarrow \Pi_j (p_j \rho_j - p_i \rho_i) \Pi_i &= 0, \forall 1 \leq i, j \leq m \end{aligned} \quad (8)$$

(8) was derived by Holevo, separately, without using the dual optimization problem. (6) and (8) are equivalent to each other. These are necessary but not sufficient conditions. The solution set contains a finite number of m-POVMs of which a proper subset is optimal. This optimal POVM will satisfy the global maxima conditions given below:

$$\begin{aligned}
Z &\geq p_i \rho_i \\
\Rightarrow \sum_{k=1}^m p_k \rho_k \Pi_k - p_i \rho_i &\geq 0, \forall 1 \leq i \leq m
\end{aligned} \tag{9}$$

Thus the necessary and sufficient conditions for the m-element POVM to globally maximize P_s is are given by (8) and (9).

3.1 Linearly Independent Pure State Ensembles

Whatever has been said till now applies to the minimum error discrimination for any ensemble of states $\{p_i > 0, \rho_i\}_{i=1}^m$. From here on we wish to examine minimum error discrimination of linearly independent pure states i.e. the ensemble is $\{p_i, |\psi_i\rangle\langle\psi_i|\}_{i=1}^m$ where $\{|\psi_i\rangle\}_{i=1}^m$ is a linearly independent set. This implies that $r_i = 1 \forall 1 \leq i \leq m$ and that $n = m$. Hence the space is now m-dimensional.

Continuous optimization problems¹ generally yield two different kinds of conditions both of which need to be solved: the first one correspond to stationary conditions of the objective function in the space of feasible solutions². These conditions are in the form of equations. The second correspond to the minimization or maximization of the objective function. These conditions are in the form of inequalities. One could adopt two different approaches for solving the problem at hand: (i) Solve both the stationary and the extrema conditions simultaneously. (ii) First solve the stationary conditions. Typically, the solution set of the stationary conditions will be discrete. Having obtained this solution set, test each solution individually for the maxima/ minima conditions.

Let's interpret what the equations (8) and (9) mean when the ensemble is that of linearly independent pure states.

It can be shown [21, 1, 6] that in this context, (8) imply that the elements of the POVM have to be rank-one. Thus the POVM comprises of m rank-one elements and covers a space that is m-dimensional. This implies that the POVM has to be a rank-one projective measurement. Thus we can restrict the candidate set to the space of rank-one projective measurements. What does (8) imply further? It can be easily shown that (8) are conditions for the stationary points of the objective function P_s in space of rank-one projective measurements. This describes fully the role played by (8) in the optimization problem.

Once (8) has been satisfied by some rank-one projective measurements, (9) is the condition for the global maximum of P_s on them. It has to be noted that this is different from the local maximum condition which wouldn't be sufficient for MED.

Solving optimization problems by (i) is more difficult than (ii), even numerically. Here we attempt to solve it analytically for the case where $n = 3$ and ensemble is real.^{3,4} We try solving the problem through both (i) and (ii).

(i.a) The MED of many general families of ensembles of qubit states can be solved using the qubit-state space geometry [10][9][11]⁵. We investigate what the problem looks like for MED for linearly independent pure states in the qutrit state space. As can be expected, the richness of the qutrit space geometry makes the problem a lot more difficult. The advantages that the qubit space geometry has to offer, owing to its simplicity and ease of visualization, are lost when it comes to the qutrit space.

¹constrained or otherwise

² The space of feasible solutions is the space which satisfies the constraints of the problem. In our case these constraints are $\Pi_i \geq 0$ and $\sum_i \Pi_i = \mathbb{1}$ i.e. the space of m element POVMs.

³A "real ensemble" implies that there exists a basis in which all states of the ensemble become real.

⁴Arguments in [21],[22] imply that when dealing with a real ensemble the conditions, equations and solutions to the problem are entirely real as well. This doesn't serve any physical purpose; just is a simpler to solve.

⁵In fact for m=2, solution is trivial

(8) and (9) give us some polynomial equations and inequalities. Given these equations and inequalities one proceeds by first solving the equations and then testing the inequalities with each element in the solution set. Thus while we started with the intention of solving by (i) we ultimately have to resort to solve by (ii). One can formulate the problem in other means as we will describe below. Those equations are much simpler to solve. The reason why we made these equations explicit is to give the general picture of how complicated the problem can become for even simplest systems.

(i.b) Starting from (8) and (9), we reformulate the problem adopting a representation in a special basis. Here we encounter some simultaneous non-linear equations. These equations aren't polynomial equations but can be manipulated to obtain polynomial equations. Yet one loses certain information through these manipulations⁶. Some of the information lost pertains to the global maxima conditions (9). Hence the new polynomial equation we obtain give us solutions for the stationary points among which one uniquely will correspond to the global maxima. Since the non-polynomial equations are more tedious to solve, it is better to opt for the polynomial equations. Once again, despite starting out to solve through (i), we resort to solving through (ii).

(ii) The polynomial equations from (i) are very complicated. One desires much more simplified equations. The representation just used implies that one can obtain close cousin of equations to those obtained in (i.b) which are much simpler to solve. For real systems (8) are cast into a set of 3 simultaneous real trivariate polynomial equations with a discrete solution set⁷. Given that \mathbb{R} isn't an algebraically closed field, this discrete solution set may include some complex solutions which are discarded. Each one of the remaining solutions is real and corresponds to a rank-one projective measurement where P_s is stationary⁸. Of these stationary points, a unique one is the global maximum whereas the others correspond to saddle points, points of inflection, local maxima, local minima and the global minimum. Each one of these solutions has to be individually tested for (9). Obtaining solutions for these polynomial equations implies solving them symbolically. As we will see later, this is a very difficult task to accomplish⁹.

3.2 Geometric Method

The geometry of the generalized Bloch sphere Ω_3 for qutrits was completely specified in [20]. Here we go through it briefly:

Any density matrix ρ has a unique representation:

$$\rho = \frac{1}{3}(\mathbb{1} + \sqrt{3}\vec{n} \cdot \vec{\lambda}) \quad (10)$$

where λ_i represent the Gell-Mann matrices and \vec{n} represent a vector from \mathbb{R}^8 which satisfy the following properties:

$$\vec{n} \cdot \vec{n} \leq 1 \quad (11)$$

$$3\vec{n} \cdot \vec{n} - 2(\vec{n} * \vec{n}) \cdot \vec{n} \leq 1 \quad (12)$$

(11) is the condition for ρ to be non-negative whereas (12)¹⁰ specifies the boundary and interior of Ω_3 .

⁶This manipulation involves squaring or cubing both sides of the equation resulting in the loss of some information.

⁷For complex systems, the number of variables and equations are six.

⁸i.e. P_s is stationary in the space of rank-one projective measurements. On the other hand, P_s cannot be stationary in the space of m-POVMs. This is because P_s is linear in the m-POVMs and space of m-POVMs is convex. Had P_s been stationary at any point in the m-POVM space, it would have to be constant for all points in the space. We know that this isn't true.

⁹It needs to be mentioned that we obtained other polynomial equations for solving (8) too. The reason we avoid mentioning them explicitly is because they involve a greater number of real equations in real variables. Suffice it to say that whatever holds for (8) holds for any such polynomial set.

¹⁰ $(\vec{n} * \vec{n})_l = \sqrt{3}d_{jkl}n_jn_k$ where $d_{jkl} = \frac{1}{4}\text{Tr}(\lambda_j\{\lambda_k, \lambda_l\})$

Let $|\psi_i\rangle\langle\psi_i|$ correspond to bloch vector $\vec{n}^{(i)}$ which saturates both (11) and (12). For $\{\vec{n}^{(i)}\}_{i=1}^3$ to correspond to a set of linearly independent vectors, we require that $\rho' = \sum p_i |\psi_i\rangle\langle\psi_i|$ lies strictly in the interior of Ω i.e. let $\vec{n} = \sum_{i=1}^3 p_i \vec{n}^{(i)}$, then $3\vec{n} \cdot \vec{n} - 2(\vec{n} * \vec{n}) \cdot \vec{n} < 1$. Here we don't impose the condition that $|\psi_i\rangle\langle\psi_i|$ be real.

From (6) and (9) we know that the operators $\tilde{\sigma}_i := Z - p_i |\psi_i\rangle\langle\psi_i|$ are positive semidefinite with kernel spanned by one dimensional $\text{Supp}(\Pi_i)$. We rewrite $\tilde{\sigma} := \kappa_i \sigma_i$ where $\text{Tr}(\sigma_i) = 1$. Hence σ_i is a rank two density operator. Since Π_i are rank 1 they are extreme points on the boundary of the bloch sphere. Let bloch vectors corresponding to σ_i be $\vec{s}^{(i)}$ and for Π_i be $\vec{t}^{(i)}$. It follows that

$$(3\vec{s}^{(i)} - 2\vec{s}^{(i)} * \vec{s}^{(i)}) \cdot \vec{s}^{(i)} = 1, \quad \vec{s}^{(i)} \cdot \vec{s}^{(i)} < 1 \quad (13)$$

$$(3\vec{t}^{(i)} - 2\vec{t}^{(i)} * \vec{t}^{(i)}) \cdot \vec{t}^{(i)} = 1, \quad \vec{t}^{(i)} \cdot \vec{t}^{(i)} = 1 \quad (14)$$

Consider

$$Z = \frac{k_0}{3} (\mathbb{1} + \sqrt{3} \vec{k} \cdot \vec{\lambda}) \quad (15)$$

where

$$\begin{aligned} \text{Max } (p_1, p_2, p_3) &\leq k_0 \leq 1 \\ \vec{k} \cdot \vec{k} &< 1 \\ (3\vec{k} - 2\vec{k} * \vec{k}) \cdot \vec{k} &\leq 1 \end{aligned} \quad (16)$$

And hence we get

$$\begin{aligned} \kappa_i &= k_0 - p_i \\ \vec{s}^{(i)} &= \frac{k_0 \vec{k} - p_i \vec{n}^{(i)}}{\kappa_i} \end{aligned} \quad (17)$$

Now $\vec{t}^{(i)}$ is a function of $\vec{s}^{(i)}$. Since $\vec{s}^{(i)}$ satisfies (13) $\vec{t}^{(i)}$ has to satisfy the following equation linear in its components:

$$\vec{s}^{(i)} + \vec{t}^{(i)} + \vec{t}^{(i)} * \vec{s}^{(i)} = \vec{0} \quad (18)$$

(13) guarantees that there is a unique solution of (18) for $\vec{t}^{(i)}$ and this solution automatically obeys (14). That said this functional dependence of $\vec{t}^{(i)}$ on $\vec{s}^{(i)}$ is extremely complicated. Now

$$\sum_i \Pi_i = \mathbb{1} \Rightarrow \sum_i \vec{t}^{(i)} = \vec{0} \quad (19)$$

The polytope formed by the points $\{p_i \vec{n}^{(i)}\}_{i=1}^3$ and $\{\kappa_i \vec{s}^{(i)}\}_{i=1}^3$ in \mathbb{R}^8 are congruent. In fact, the latter is a displaced mirror image of the former.

$$p_i - p_j = \kappa_j - \kappa_i \quad (20)$$

$$p_i \vec{n}^{(i)} - p_j \vec{n}^{(j)} = \kappa_j \vec{s}^{(j)} - \kappa_i \vec{s}^{(i)} \quad (21)$$

We also require that the identity should be resolved by the states orthogonal to σ_i (i.e. Π_i). One of the main hurdles faced in the qutrit case is to **visualize** where $\vec{t}^{(i)}$ should be located so that $\text{Tr}(\sigma_i \Pi_i) = 0$

is satisfied. Also, one needs to satisfy (19). In the qubit case σ_i would have been a rank-one pure state implying that $\vec{s}^{(i)}$ would be on the surface of the bloch sphere and $\vec{t}^{(i)}$ would be the former's corresponding anti-podal point. These simplicities are no more in the qutrit case. In general there are multiple set of states $\{\sigma_i\}_{i=1}^3$ for which some κ_i can be found so that (20) and (21) are satisfied. The problem now is to ensure that (19) too is satisfied. The only way one can proceed is to solve the the system of 9 unknowns k_0, \vec{k} for (16),(13),(18) and (19), where $\vec{s}^{(i)}$ is given by (17), algebraically. eqrefsb,(18) and (19) are all polynomial equations whereas (16) are inequalities. Satisfying these equations gives us the optimal POVM $\vec{t}^{(i)} \rightarrow \Pi_i$. Explicit form of $\vec{t}^{(i)}$ in terms of components of $\vec{s}^{(i)}$ make the polynomial equations really tedious. It's desirable to obtain a much simpler set of equations. With that in mind we turn to a new method.

3.3 Gram Matrix Method

We first talk about the general n-dimensional systems. Later on we specialize for n=3.

We wish to obtain the optimal POVM (which is a rank-one projective measurement) for MED for an ensemble $\{p_i, |\psi_i\rangle\langle\psi_i|\}_{i=1}^m$ where $\{|\psi_i\rangle\}_{i=1}^m$ is a linearly independent set. Let $|\tilde{\psi}_i\rangle = \sqrt{p_i}|\psi_i\rangle, \forall 1 \leq i \leq m$. Since $\{|\tilde{\psi}_i\rangle\}_{i=1}^m$ form a linearly independent set, there exists a corresponding unique ordered set $\{|\tilde{u}_i\rangle\}_{i=1}^m$ where $|\tilde{u}_i\rangle \in \mathcal{H}$ such that

$$\langle\tilde{\psi}_i|\tilde{u}_j\rangle = \delta_{i,j}, \forall 1 \leq i, j \leq m \quad (22)$$

Let G denote the gram matrix of $\{|\tilde{\psi}_i\rangle\}_{i=1}^m$. The matrix elements of G are hence given by

$$G_{ij} = \langle\tilde{\psi}_i|\tilde{\psi}_j\rangle, \forall 1 \leq i, j \leq m \quad (23)$$

Since $\{|\tilde{\psi}_i\rangle\}_{i=1}^m$ is linearly independent set, $G \geq 0$. The gram matrix of $\{|\tilde{u}_i\rangle\}_{i=1}^m$ is G^{-1} . Any ordered orthonormal basis $\{|v_i\rangle\}_{i=1}^m$ can be represented as follows:

$$|v_i\rangle = \sum_{j=1}^m (G^{\frac{1}{2}}U)_{ji} |\tilde{u}_j\rangle \quad (24)$$

where $G^{\frac{1}{2}}$ is the positive square root of G and U is an n dimensional unitary matrix. U captures the unitary degree of freedom of the ordered orthonormal basis. One can easily check that $\langle v_i | v_j \rangle = \delta_{i,j}$. Any such an ordered orthonormal basis corresponds to an n-element rank-one projective measurement:

$$\{|v_i\rangle\}_{i=1}^m \rightarrow \{|v_i\rangle\langle v_i|\}_{i=1}^m \quad (25)$$

In particular, when $U = \mathbb{1}$ we obtain the pretty good measurement associated with $\{|\tilde{\psi}_i\rangle\}_{i=1}^m$. Now consider that one can change the phase factors of the ONB vectors by appending a diagonal unitary, U' on the right of U in (24) where $U'_{jk} = \delta_{j,k} e^{i\theta_j}$. This implies $\{|v_j\rangle\}_{j=1}^m \rightarrow \{e^{i\theta_j} |v_j\rangle\}_{j=1}^m$. Changing these phase factors doesn't change the rank-one projective measurement the basis corresponds to. If one were to use this rank-one projective measurement for MED, the probability of success on average is given by:

$$P_s = \sum_{i=1}^n |\langle\tilde{\psi}_i|v_i\rangle|^2 = \sum_{i=1}^n |(G^{\frac{1}{2}}U)_{ii}|^2 \quad (26)$$

The terms $|(G^{\frac{1}{2}}U)_{ij}|^2$ where $i \neq j$ is the error probability that B's measurement yields j despite A sending him ρ_i . Note that when U is unity, the probability of ith POVM-element clicking while ρ_j was sent is equal to the probability of jth POVM-element clicking while ρ_i was sent.

Now using (24) in (8) we get:

$$(G^{\frac{1}{2}}U)_{jj}(G^{\frac{1}{2}}U)_{jk}^* = (G^{\frac{1}{2}}U)_{kj}(G^{\frac{1}{2}}U)_{kk}^* \quad \forall 1 \leq j, k \leq m \quad (27)$$

which is a condition on U .

For the optimal POVM $\{|\tilde{v}_i\rangle\langle\tilde{v}_i|\}_{i=1}^m$, Carlos Mochon [21] showed that

$$|\langle\tilde{\psi}_i|\tilde{v}_i\rangle|^2 > 0 \quad \forall 1 \leq i \leq m \quad (28)$$

Suppose \tilde{U} corresponds to the optimal rank one projective measurement through the correspondence given by (25). Then (28) implies that the diagonal elements of $G^{\frac{1}{2}}\tilde{U}$ have to be non-zero. Also one can use U' to make these diagonal elements positive. Thus (27) becomes:

$$(G^{\frac{1}{2}}\tilde{U})_{jj}(G^{\frac{1}{2}}\tilde{U})_{jk}^* = (G^{\frac{1}{2}}\tilde{U})_{kj}(G^{\frac{1}{2}}\tilde{U})_{kk} \quad \forall 1 \leq j, k \leq m \quad (29)$$

Now we have the matrix equation:

$$(\tilde{U}^\dagger G^{\frac{1}{2}})G^{-1}(G^{\frac{1}{2}}\tilde{U}) = \mathbb{1} \quad (30)$$

Let D be a positive diagonal matrix comprising the diagonals of $G^{\frac{1}{2}}\tilde{U}$:

$$D_{i,j} = \delta_{i,j}(G^{\frac{1}{2}}\tilde{U})_{i,i} \quad \forall 1 \leq i, j \leq m \quad (31)$$

From (29) we infer that the matrix $DG^{\frac{1}{2}}\tilde{U}$ is hermitian. Also since $D > 0$, D^{-1} exists. From (30) we then get:

$$(\tilde{U}^\dagger G^{\frac{1}{2}}D)(DGD)^{-1}(D G^{\frac{1}{2}}\tilde{U}) = \mathbb{1} \quad (32)$$

From (32) we can see that $DG^{\frac{1}{2}}\tilde{U}$ is a hermitian square root of the matrix DGD . If we represent $D_{ii} = a_i \quad \forall 1 \leq i \leq m$ we see that the diagonal elements of $DG^{\frac{1}{2}}\tilde{U}$ are a_i^2 . Thus we can solve the stationary condition (29) by finding a set of positive real numbers $\{a_i\}_{i=1}^m$ such that a hermitian square root of the matrix DGD has diagonal entries equal to a_i^2 . Having obtained such a set of positive numbers, one can proceed to find the associated rank-one projective measurement which will satisfy the stationary conditions (8).

For a given G there are many such sets of positive reals $\{a_i\}_{i=1}^m$ which satisfy this desired property. Each of these different sets correspond to different rank-one projective measurements each of which will satisfy the stationary condition (8). In keeping with the fact that only one POVM can be optimal, only one of these rank-one projective measurements satisfy the global maxima condition (9). Carlos Mochon and Belavkin [21][22] noted that this optimal solution corresponds to the case where $DG^{\frac{1}{2}}\tilde{U}$ is the positive square root of DGD . The corresponding projective measurement is the pretty good measurement for another ensemble with the same states $\{|\psi_i\rangle\}_{i=1}^m$ but with different probabilities.

3.3.1

Here we obtain algebraic equations to solve for the set $\{a_1, a_2, a_3\}$ for the case $m = 3$ and where the ensemble is real. G is given to us. a_1, a_2, a_3 are our unknowns. The equations are obtained as follows:

$$(\sqrt{DGD})_{ii} = a_i^2 \quad \forall 1 \leq i \leq 3 \quad (33)$$

The expressions for $(\sqrt{\text{DGD}})_{ii}$ in terms of the matrix elements of G and in terms of the unknowns a_1, a_2, a_3 are extremely complicated. Here we do not write the full expression down, but symbolically denote matrix elements of $(\sqrt{\text{DGD}})_{ii}$ as:

$$(\sqrt{\text{DGD}})_{ii} = \sum_j^3 \sqrt{\lambda^{(j)}} |\zeta^{(j)}_i|^2 \quad (34)$$

where $\lambda^{(j)}$ are the eigenvalues of DGD , $\vec{\zeta}^{(j)}$ the corresponding eigenvector and $\zeta^{(j)}_i$ the i th component of this eigenvector. Now $\lambda^{(j)}$ and $\zeta^{(j)}_i$ are complicated functions of the unknowns a_1, a_2 and a_3 and matrix elements of G .¹¹ The equations (33) subsume not only (8) but the global maxima conditions (9).

Given the complicated expression of the equations in a_1, a_2, a_3 , one is discouraged from solving these simultaneous equations analytically. It would be better to undertake the tedious exercise to cast them into polynomial equations by rearranging terms in the equations and squaring and cubing the equations. This necessarily means that the information of $(\sqrt{\text{DGD}})_{ii}$ corresponding to matrix elements of the **positive** square root of DGD is lost. Thus solutions from the resulting polynomial equations will always solve (8) but only one of them will solve (9).

3.3.2

In this subsection we obtain a different set of simultaneous multivariate polynomial equations. These are more easily obtained. Just that the number of unknowns increases to more than m . To reduce the number of unknowns we restrict ourselves to the case where $m = 3$ and where the gram matrix G is real.

From (29) we infer that the matrix $G^{\frac{1}{2}} \tilde{U} D^{-1}$ is hermitian with unit diagonal elements.

From (30) we get

$$(D^{-1} \tilde{U}^\dagger G^{\frac{1}{2}}) G^{-1} (G^{\frac{1}{2}} \tilde{U} D^{-1}) = D^{-2} \quad (35)$$

G being real implies that G^{-1} is also real. In that case the matrix $G^{\frac{1}{2}} \tilde{U} D^{-1}$ must be real symmetric with unit diagonal elements.

Let

$$G^{\frac{1}{2}} \tilde{U} D^{-1} = \begin{pmatrix} 1 & \alpha & \beta \\ \alpha & 1 & \gamma \\ \beta & \gamma & 1 \end{pmatrix} \quad (36)$$

where α, β and γ are real numbers. Then (35) becomes:

$$\begin{aligned} & \begin{pmatrix} 1 & \alpha & \beta \\ \alpha & 1 & \gamma \\ \beta & \gamma & 1 \end{pmatrix} \begin{pmatrix} (G^{-1})_{11} & (G^{-1})_{12} & (G^{-1})_{13} \\ (G^{-1})_{21} & (G^{-1})_{22} & (G^{-1})_{23} \\ (G^{-1})_{31} & (G^{-1})_{32} & (G^{-1})_{33} \end{pmatrix} \begin{pmatrix} 1 & \alpha & \beta \\ \alpha & 1 & \gamma \\ \beta & \gamma & 1 \end{pmatrix} \\ &= \begin{pmatrix} (D_{11})^{-2} & 0 & 0 \\ 0 & (D_{22})^{-2} & 0 \\ 0 & 0 & (D_{33})^{-2} \end{pmatrix} \end{aligned} \quad (37)$$

Here we get 6 different equations:

¹¹In [21] an invertible mapping is made from and to the space of non-zero m probabilities in the form: $p_i = C \frac{\tilde{p}_i}{(\tilde{G}^{\frac{1}{2}})_{ii}}$ where $1 \leq i \leq m$. Then (34) is precisely the exercise of obtaining the inverse function.

$$\begin{aligned} \alpha^2(G^{-1})_{12} + \alpha((G^{-1})_{11} + (G^{-1})_{22} + (G^{-1})_{13}\beta + (G^{-1})_{23}\gamma) \\ + (G^{-1})_{33}\beta\gamma + (G^{-1})_{23}\beta + (G^{-1})_{13}\gamma + (G^{-1})_{12} = 0 \end{aligned} \quad (38)$$

$$\begin{aligned} \beta^2(G^{-1})_{13} + \beta((G^{-1})_{11} + (G^{-1})_{33} + (G^{-1})_{23}\gamma + (G^{-1})_{12}\alpha) \\ + (G^{-1})_{22}\alpha\gamma + (G^{-1})_{12}\gamma + (G^{-1})_{23}\alpha + (G^{-1})_{13} = 0 \end{aligned} \quad (39)$$

$$\begin{aligned} \gamma^2(G^{-1})_{23} + \gamma((G^{-1})_{22} + (G^{-1})_{33} + (G^{-1})_{13}\beta + (G^{-1})_{12}\alpha) \\ + (G^{-1})_{11}\alpha\beta + (G^{-1})_{12}\beta + (G^{-1})_{13}\alpha + (G^{-1})_{23} = 0 \end{aligned} \quad (40)$$

$$\begin{aligned} (G^{-1})_{22}\alpha^2 + (G^{-1})_{33}\beta^2 + 2\alpha\beta(G^{-1})_{23} + 2\alpha(G^{-1})_{12} + 2\beta(G^{-1})_{13} \\ + (G^{-1})_{11} = (D_{11}^{-2}) \end{aligned} \quad (41)$$

$$\begin{aligned} (G^{-1})_{11}\alpha^2 + (G^{-1})_{33}\gamma^2 + 2\alpha\gamma(G^{-1})_{13} + 2\alpha(G^{-1})_{12} + 2\gamma(G^{-1})_{23} \\ + (G^{-1})_{22} = (D_{22}^{-2}) \end{aligned} \quad (42)$$

$$\begin{aligned} (G^{-1})_{11}\beta^2 + (G^{-1})_{22}\gamma^2 + 2\beta\gamma(G^{-1})_{12} + 2\beta(G^{-1})_{13} + 2\gamma(G^{-1})_{23} \\ + (G^{-1})_{33} = (D_{33}^{-2}) \end{aligned} \quad (43)$$

We obtain solutions for (α, β, γ) from (38), (39) and (40) and then substitute them in the next three equations (41), (42) and (43) to obtain the values of D_{ii} from which we can now find out what $G^{\frac{1}{2}}\tilde{U}$ is and from there obtain the optimal projective measurement.

One obtains 8 solutions for the equations (38), (39) and (40) for (α, β, γ) . In certain instances some of the solutions in the solution set aren't real. By substituting these complex (α, β, γ) in (36) we obtain a complex symmetric matrix which isn't hermitian as was required. Hence such complex solutions are unphysical in the sense that they don't correspond to some rank-one projective measurement. All the remaining solutions are real and correspond to some rank-one projective measurement. The uniqueness of the optimal POVM for linearly independent pure state ensembles implies that only one of these real solutions corresponds to the optimal POVM. The physical significance of the other real solutions is that they provide those points in the space of rank-one projective measurements which are stationary for the function P_s as given by (2). How does one identify the solution corresponding to the optimal POVM? From the previous section we know that the matrix $DG^{\frac{1}{2}}\tilde{U}$ has to be a positive matrix to correspond to the optimal POVM. Since $DG^{\frac{1}{2}}\tilde{U}$ and $G^{\frac{1}{2}}\tilde{U}D^{-1}$ are related by a congruence transformation, the optimal solution in our case must correspond to those values of (α, β, γ) for which the matrix $G^{\frac{1}{2}}\tilde{U}D^{-1}$ is positive definite. The uniqueness of the optimal POVM then implies that there's only one solution of (α, β, γ) for which the matrix $G^{\frac{1}{2}}\tilde{U}D^{-1}$ is positive definite.

Symbolically the reduced Groebner basis of the ideal of these polynomials contains 14 different elements. "Solving" the system would imply that one has to divide the parameter space into disjoint "cells" and specify a regular chain for each such cell[25]. In doing even this, one cannot be sure if the resulting polynomials will be of degree ≤ 4 in the main variable in each polynomial (with respect to some suitable monomial ordering criteria). Hence we see that even by restricting ourselves to $n = 3$ and a real gram matrix, obtaining a closed form solution for the optimal POVM is still very tough.

4 Analytic Properties of Optimal POVM for Pure State Ensembles

In this section we analyze the optimal m-POVM for MED of m linearly independent pure states as a function of the ensemble. To do that we first describe this space of m -element ensembles. We denote it by \mathfrak{E} and a point \mathcal{E} in \mathfrak{E} is of the form:

$$\mathcal{E} = \{p_i > 0, |\psi_i\rangle\langle\psi_i|\}_{i=1}^m \quad (44)$$

Consider the space of pure state ensembles. When elements in \mathcal{E} are indexed, \mathcal{E} can correspond to it a gram matrix, G . This gram matrix should be $m \times m$, positive definite and have trace 1. Changing the index-sequence of elements in \mathcal{E} is equivalent to performing a permutation transformation on the gram matrix. Additionally, by changing the phases associated with the pure states in \mathcal{E} - $|\psi_j\rangle \rightarrow e^{i\phi_j}|\psi_j\rangle$, the gram matrix transforms under a diagonal unitary transformation. Varying these phases doesn't change the state $p_i|\psi_i\rangle\langle\psi_i|$ and doesn't change the ensemble. Yet the gram matrix changes. Thus, for the same ensemble, the corresponding gram matrix can vary upto a complex permutation. We want a more faithful representation so we define an equivalence class on the set of $m \times m$ positive definite matrices with trace 1¹². We denote the quotient space by \mathfrak{G} . To represent elements in \mathfrak{G} we adopt two conventions - (i) (ordering convention) $p_i \geq p_{i+1}$; in case of equality, $|G_{i \ i+1}| \geq |G_{i+1 \ i+2}|$ and so on and so forth and (ii) (phase convention, after satisfying the ordering convention) $G_{i \ i+1} = G_{i+1 \ i} \ \forall 1 \leq i \leq m$.

$$\mathfrak{G} = \{G \text{ is } m \times m \mid G > 0, \text{Tr}(G) = 1; G \text{ obeys ordering and phase conventions.}\} \quad (45)$$

\mathfrak{G} is a convex set. This fact will come in handy later on.

$\forall \mathcal{E} \in \mathfrak{E}^{13}$, \exists a unique gram matrix, $G \in \mathfrak{G}$ corresponding to it. Let this mapping be represented as: $\mathcal{G}(\mathcal{E}) = G$.

Consider a trajectory in the space \mathfrak{E} :

$$E : [0, 1] \longrightarrow \mathfrak{E} \quad (46)$$

Thus $E(t)$ represents a point in \mathfrak{E} for $0 \leq t \leq 1$. The function E is well behaved with respect to the variable t ¹⁴. How does P_s^{max} vary along this trajectory? From the definition of P_s , i.e. (2), we see that it is continuous both in the ensemble input $\{p_i, |\psi_i\rangle\langle\psi_i|\}$ and the POVM input $\{\Pi_i\}$. If P_s^{max} were to vary discontinuously as t is varied from 0 to 1, the behaviour of P_s would have to suffer discontinuities as a function of t too. Hence P_s^{max} varies continuously as t varies from 0 to 1¹⁵. Corresponding to the map $E(t)$ we get a similar map: $G : [0, 1] \longrightarrow \mathfrak{G}$ where $G(t) = \mathcal{G}(E(t))$.

The the set of m -element POVMs forms a convex set. A POVM is generally an ordered set of m positive operators which sum to the $\mathbb{1}$. Convex sum of two POVMs is given by point-wise addition of i th elements: - $\{V_i^{(1)}\}_{i=1}^m$ and $\{V_i^{(2)}\}_{i=1}^m$, the i th element of the new POVM $\{V_i\}_{i=1}^m$ is the convex sum of respective i th elements from $\{V_i^{(1)}\}_{i=1}^m$ and $\{V_i^{(2)}\}_{i=1}^m$ i.e. $V_i = pV_i^{(1)} + (1-p)V_i^{(2)}$.

Suppose $\{V_i^{(1)}\}_{i=1}^m$ and $\{V_i^{(2)}\}_{i=1}^m$ are projective measurements. Any convex combination of these projective measurements is a projective measurement iff either if (i) $p = 0$ (or $p = 1$) OR (ii) $\{V_i^{(1)}\}_{i=1}^m = \{V_i^{(2)}\}_{i=1}^m$ ¹⁶.

The optimal POVM for linearly independent pure states *has to be* an m -element projective measurement. This implies that the optimal POVM *has to be* unique.

Now as t varies from 0 to 1 how does the optimal POVM vary? The optimal POVM will be an m -element rank one projective measurement, so we can confine our focus on this space itself. The first thing to notice is that the uniqueness of the optimal POVM implies that we, actually, *can* define a mapping $V : [0, 1] \longrightarrow \mathfrak{V}$ such that $V(t)$ is the optimal POVM for MED of $E(t)$. Is $V(t)$ continuous in t ¹⁷? Suppose not; let there be a jump at some point t_0 so that $\lim_{t \xrightarrow{\text{one-side}} t_0} V(t) \neq V(t_0)$. Since P_s is continuous function in both its ensemble input and its POVM input, we require that $P_s^{max}(t_0) > \lim_{t \rightarrow t_0} P_s^{max}(t)$: a contradiction. Thus $V(t)$ has to vary continuously as a function of t .

¹²Two matrices from this set are equivalent iff one is a complex permutation of other other.

¹³Upto a unitary transformation of its elements: $|\psi_i\rangle\langle\psi_i| \rightarrow U|\psi_i\rangle\langle\psi_i|U^\dagger \ \forall 1 \leq i \leq m$

¹⁴i.e. the a priori probabilities and matrix elements of all pure states in $E(t)$ are analytic functions of t

¹⁵This holds regardless of whether the ensemble is linearly independent or not.

¹⁶Both **ordered sets** have to be equal. Also, this holds even for projective measurements which aren't rank 1.

¹⁷i.e. matrix elements of $V(t)$ continuous in t

From (32) we know that $V(t)$ has to be the pretty good measurement associated with some ensemble of the same states but different a priori probability. At the gram matix level, both ensembles are related by $G \longrightarrow \frac{1}{\text{Tr}(DGD)} DGD$ where D is a positive diagonal matrix such that the positive \sqrt{DGD} (which is positive definite) has diagonal elements a_{ii}^2 (see (3.3.1)).

We can construct $V(t)$ from $\sqrt{D(t) G(t) D(t)}$ using simple arithmetic operations of the matrix elements of the latter(see (3.3)). Thus, if $D(t)$ and $D(t) G(t) D(t)$ vary analytically as a function of t , then so does $V(t)$. To inspect if the former is true, we employ the implicit function theorem.

Implicit Function Theorem: Let $\{y_i\}_{i=1}^N$ be N functions (real or complex) of the independent variables - $\{t, f_1, f_2, \dots, f_N\}$ where the variables $\{f_i\}_i$, which are N in number too. Let $(\tau, \phi_1, \phi_2, \dots, \phi_N)$ be a point such that $y_i(\tau, \phi_1, \phi_2, \dots, \phi_N) = 0 \ \forall \ 1 \leq i \leq N$. If the Jacobian matrix $J_{i,j} = \frac{\partial y_i}{\partial f_j}$ is invertible at $(\tau, \phi_1, \phi_2, \dots, \phi_N)$ then there exists some open neighbourhood, T , containing τ : for which there exist open neighbourhoods S_i containing ϕ_i such that $f_i : T \longrightarrow S_i$ can be defined, so that $y_i(t, f_1(t), f_2(t), \dots, f_N(t)) = 0 \ \forall \ 1 \leq i \leq N, \ \forall t \in T$. That is

$$\{(t, \vec{f}) \in T \times S \mid \vec{y}(t, \vec{f}) = 0\} = \{(t, \vec{f}(t)) \mid t \in T \text{ and } y(t, \vec{f}(t)) = 0\} \quad (47)$$

where $S = S_1 \times S_2 \times \dots \times S_N$.

Simply put, the implicit function theorem gives sufficient conditions for the variables f_i to implicitly depend on the variable t near some point (τ, ϕ_i) so that in some neighbourhood of this point $\vec{y}(t, \vec{f}(t))$ is constant in t .

Analytic Implicit Function Theorem: Furthermore if y_i is an analytic function in the variables f_i then the implicit dependence of f_i on t will be analytic too.

From (32) we get that:

$$(DG^{\frac{1}{2}}\tilde{U})^2 - DGD = 0 \quad (48)$$

where $DG^{\frac{1}{2}}\tilde{U}$ is the positive square root of DGD and $(DG^{\frac{1}{2}}\tilde{U})_{ii} = a_{ii}^2$. In our application of IFT, t is the independent variable. The implicit variables f_{ij} and a_i are defined as:

$$f_{ij}(t) \longrightarrow (DG^{\frac{1}{2}}\tilde{U})_{ij} \text{ when } i \neq j \quad (49)$$

$$a_i(t) \longrightarrow \sqrt{(DG^{\frac{1}{2}}\tilde{U})_{ii}} \quad (50)$$

The implicit variables $\{a_i, f_{ij} | i \neq j\}$ are m^2 in number. Here these variables are complex, thus f_{ij} and f_{ji} are explicitly taken different.

Define $F(t)$ to be an $m \times m$ matrix defined by:

$$F_{ij}(t) = f_{ij}(t) \text{ when } i \neq j \quad (51)$$

$$F_{ii}(t) = a_i(t)^2 \quad (52)$$

Consider the matrix equation:

$$(F(t))^2 - D(t)G(t)D(t) = 0 \quad (53)$$

This the same as (48).

Define a new matrix $Y(t)$ as:

$$y_{ij}(t) = ((F(t))^2 - D(t)G(t)D(t))_{ij} \quad (54)$$

where $(Y(t))_{ij} = y_{ij}(t)$.

Hence, we want the functions $\{a_i(t), f_{ij}(t)\}$ to vary implicitly on t in a manner such that the functions $y_{ij}(t) = 0$.

The Jacobian is given by

$$J_{(ij),(kl)} = \frac{\partial y_{ij}}{\partial f_{kl}} \text{ where } k \neq l \quad (55)$$

$$J_{(ij),(kk)} = \frac{\partial y_{ij}}{\partial a_k} \quad (56)$$

The Jacobian matrix is an $m^2 \times m^2$ matrix. Now a matrix is invertible iff it's rows (or columns) are linearly independent. We de-vectorize the rows of the Jacobian matrix and bring them in the form of $m \times m$ matrices. If *these* matrices (which are m^2 in number) are linearly independent then that proves that the Jacobian is invertible.

Define matrix $M^{i,j}$ by:

$$M_{kl}^{ij} = J_{(ij),(kl)} \quad (57)$$

This matrix is of the form:

$$M_{ij} = \begin{matrix} & & & \text{(j}^{\text{th}} \text{ column)} \\ & & & \downarrow \\ \begin{matrix} \text{(i}^{\text{th}} \text{ row} \rightarrow) \end{matrix} & \begin{pmatrix} 0 & 0 & \cdots & X & \cdots & 0 \\ 0 & 0 & \cdots & X & \cdots & 0 \\ \vdots & \vdots & \vdots & X & \vdots & 0 \\ X & X & X & X & X & X \\ 0 & 0 & \cdots & X & \cdots & 0 \\ \vdots & \vdots & \vdots & X & \ddots & 0 \\ 0 & 0 & \cdots & X & \cdots & 0 \end{pmatrix} \end{matrix} \quad (58)$$

where the X 's are polynomials in the implicit variables $\{a_i(t), f_{ij}(t)\}$ and matrix elements of $G(t)$. While it is reasonable to expect that the matrices M^{ij} are linearly independent (implying the the Jacobian is invertible) unless we know how the implicit variables behave as functions of t , we can't conclude anything about their linear independence for all points in \mathfrak{G} . This puts us in a cyclic conundrum. The implicit function theorem provides a sufficient but not necessary condition for the existence of such implicit functions. That is, there are pathological examples where an implicit function exists even at points where Jacobian is singular. In this case, our inability to establish that the Jacobian isn't singular at all points $G \in \mathfrak{G}$ isn't a cause for worry. This is because we know beforehand [21][22] that such an implicit function exists, that it is continuous and furthermore we know that it is globally one-to-one with respect $G \in \mathfrak{G}$.

Implicit dependence does not exist when a critical point is a bifurcation point. In our case we know that such a point cannot exist. Otherwise, we'd have two different optimal POVM for the same point in \mathfrak{G} . And we know that this cannot happen for linearly independent ensembles.

Now that we have dealt with the question of the invertibility of the Jacobian we note that the y_{ij} are analytic functions of the variables $t; a_i, f_{ij}$. Thus $a_i(t), f_{ij}(t)$ have to be analytic in terms of t ¹⁸. This

¹⁸From the analytic implicit function theorem.

piece of information is important because it tells us that the diagonal matrix $D(t)$ varies analytically with t . And this, in turn implies, as we argued before, that the optimal POVM will vary analytically with the ensemble $V(t)$.

4.1 Analytic Continuation of the Optimal POVM

Having established that the optimal POVM will vary analytically as a function of the ensemble at all points $G \in \mathfrak{G}$, we ask the following question: Can one **drag** the optimal POVM from one ensemble to another by employing some technique?

For this, first of all, we need a trajectory. Using the mapping \mathcal{G} , one can find a simple trajectory between two ensembles. Let E_1 represent the first ensemble and $G_1 = \mathcal{G}(E_1)$. Let the optimum POVM for E_1 be known. That is to say, we know what D_1 is. Let E_2 be another ensemble and $G_2 = \mathcal{G}(E_2)$. We want to obtain D_2 . Since \mathfrak{G} is a convex set, one can obtain a simple linear trajectory between them:

$$G(t) = (1 - t) G_1 + t G_2 \text{ where } t \in [0, 1] \quad (59)$$

For any value of $t \in [0, 1]$, $G(t) \in \mathfrak{G}$.

We've established the trajectory. Next we need some differential equations which we use to drag $D(t)$ from D_1 to D_2 . For this employ IFT again: take the total derivative w.r.t of the equation (54) for all matrix elements and set it equal to zero.

$$\frac{d y_{kl}}{d t} = \zeta_{kl}(t; a_i(t), f_{ij}(t), \frac{da_i(t)}{dt}, \frac{df_{ij}(t)}{dt}) = 0 \quad \forall 1 \leq k, l \leq m \quad (60)$$

This gives us a set of m^2 first order coupled ordinary differential equations for the implicit variables $a_i(t)$, $f_{ij}(t)$.

The form of ζ_{kl} in terms of a_i , f_{ij} , $\frac{da_i}{dt}$, $\frac{df_{ij}}{dt}$ is rather complicated and depends on the value of m . As an illustration we write down what ζ_{kl} look like for $m = 2$:

$$\begin{aligned} \zeta_{11} &= 4a_1(t)^3 \frac{da_1(t)}{dt} + f_{12}(t) \frac{df_{21}(t)}{dt} + f_{21}(t) \frac{df_{12}(t)}{dt} - 2a_1(t)g_{11}(t) \frac{da_1(t)}{dt} - a_1(t)^2 \frac{dg_{11}(t)}{dt} \\ &= 0 \end{aligned} \quad (61)$$

$$\begin{aligned} \zeta_{12} &= (a_1(t)^2 + a_2(t)^2) \frac{df_{12}(t)}{dt} + (2a_1(t)f_{12}(t) - a_2(t)g_{12}(t)) \frac{da_1(t)}{dt} + \\ &\quad (2a_2(t)f_{12}(t) - a_1(t)g_{12}(t)) \frac{da_2(t)}{dt} - a_1(t)a_2(t) \frac{dg_{12}(t)}{dt} \\ &= 0 \end{aligned} \quad (62)$$

$$\begin{aligned} \zeta_{21} &= (a_1(t)^2 + a_2(t)^2) \frac{df_{21}(t)}{dt} + (2a_1(t)f_{21}(t) - a_2(t)g_{21}(t)) \frac{da_1(t)}{dt} \\ &\quad + (2a_2(t)f_{21}(t) - a_1(t)g_{21}(t)) \frac{da_2(t)}{dt} - a_1(t)a_2(t) \frac{dg_{21}(t)}{dt} \\ &= 0 \end{aligned} \quad (63)$$

$$\begin{aligned} \zeta_{22} &= 4a_2(t)^3 \frac{da_2(t)}{dt} + f_{12}(t) \frac{df_{21}(t)}{dt} + f_{21}(t) \frac{df_{12}(t)}{dt} - 2a_2(t)g_{22}(t) \frac{da_2(t)}{dt} - a_2(t)^2 \frac{dg_{22}(t)}{dt} \\ &= 0 \end{aligned} \quad (64)$$

Here $g_{ij}(t) = G(t)_{ij}$ is where the explicit time dependence comes into the equations.

These equations become more complicated as m increases. These equations are hermiticity preserving i.e. $a_i(t) \in \mathbb{R} \Rightarrow a_i(t + \delta) \in \mathbb{R}$ and $f_{ij}(t) = f_{ji}(t)^* \Rightarrow f_{ij}(t + \delta) = f_{ji}(t + \delta)^*$ for infinitesimal δ .

The equations give us a way of obtaining the optimal POVM by dragging the solution from a point where the solution is known to our desired point where the solution is unknown. It is worth mentioning how everything fits into the technique. We start from a point whose solution we know and we drag the

solution from one point to another using our trajectory. This dragging is based on the rule that (29) (and hence (8)) is always satisfied, i.e. if at a point t we have $(a_i, f_{ij}) \ni (29)$ is satisfied we ask what values for (a_i, f_{ij}) will satisfy (29) at the point $t + \delta$. Note that (29), by itself, is merely a necessary condition. (9) needs to also be satisfied for the required solution. Now $(9) \iff F(t) > 0$. How do we know that given a trajectory within \mathfrak{G} the differential equations will preserve the positivity of the matrix $F(t)$? $F(0)$ is positive definite. Let's suppose there is a point, t_1 where one of the eigenvalues of $F(t_1)$ is negative. That means that there must have been some previous point (say, t_0 at which this particular eigenvalue would have been 0. From (53) it's implied that the matrix $D(t_0)$ must also be singular at this point. From $t=0$ to $t=t_0$, $F(t) > 0$ and $D(t) > 0$. This implies that for all ensembles from $t=0$ to $t=t_0$, solution has been obtained. Examine that as we approach the limit $t \xrightarrow{\text{from } 0} t_0$, one of the values of $a_i \rightarrow 0$. Such scenarios can only arise when (i) $p_i(t) \rightarrow 0$. Then at $t = t_0$, we just have a set of $m-1$ states in our ensemble. OR (ii) The part of $|\psi_i(t)\rangle$ that is perpendicular to other states in the ensembles tends to zero, i.e. $\lim_{(t-t_0)^-} |u_i(t)\rangle = 0$. In either case $G(t_0) \notin \mathfrak{G}$ which implies that $G(t)$ isn't continuous in \mathfrak{G} . Hence we can conclude that as long as the trajectory is safely within \mathfrak{G} , we won't encounter this problem.

We employed this method to obtain solutions for various linearly independent pure state ensembles for cases: $m = 2, 3, 4$ and 5 . In all the different cases our starting point was the equiprobable orthogonal ensemble for which the solution is trivial i.e. $G_1 = \frac{1}{m} \mathbb{1}$. We employed the RK4 method to solve for many unknown ensembles (generated randomly). And in all cases we managed to construct a POVM which satisfied (8) and (9) i.e. we constructed the optimal POVM for this case.

One can obtain the solution with as much precision as desired using RK4. For RK4, the local truncation varies as $\mathcal{O}(h^5)$ and the total accumulated error as $\mathcal{O}(h^4)$. Our method gives us a way of measuring the error without having to compare results with other techniques [23][24][5]. Consider (53). The deviation of the RHS from 0 is an indication of how much error has seeped in. Thus we plot Hilbert-Schmidt norm of the RHS of (53) as a function of the iteration to get an idea of how much error accumulates. We illustrate with an example when $m = 5$. Let the gram matrix of the ensemble be given by:

$$G = \begin{pmatrix} 0.3 & \sqrt{0.06}(0.2 + i0.1) & \sqrt{0.06}(0.1) & \sqrt{0.045}(0.1) & \sqrt{0.045}(0.1) \\ \sqrt{0.06}(0.2 - i0.1) & 0.2 & 0.06 & \sqrt{0.03}(0.2 + i0.2) & \sqrt{0.03}(0.1) \\ \sqrt{0.06}(0.1) & 0.06 & 0.2 & \sqrt{0.03}(0.2 + i0.05) & \sqrt{0.03}(0.3 + i0.2) \\ \sqrt{0.045}(0.1) & (0.2 - i0.2)\sqrt{0.03} & \sqrt{0.03}(0.1 - i0.05) & 0.15 & (0.15)(0.2 + i0.3) \\ \sqrt{0.045}(0.1) & (0.1)\sqrt{0.03} & (0.3 - i0.3)\sqrt{0.03} & (0.15)(0.2 - i0.3) & 0.15 \end{pmatrix}$$

The interval length is given by $h = 10^{-3}$ and the number of iterations is 1000. We plot the log of the Hilbert Schmidt norm of the RHS of (53).

From the figure we see that for the first few iterations the truncation error is of order -16.8 and over the last few iterations total accumulated error is of order -15.6 which is well within the error-performance given by RK4. For many randomly selected ensembles the local truncation error is of order -16 and total accumulated error is of order -15 - again well within the RK4 error margin. This shows that this method is a reliable method to obtain the optimal POVM for linearly independent pure states.

Where does the method fail? If the ensemble is nearly linearly dependent (i.e. $\exists i \ni |\langle \psi_i | u_i \rangle|$ is *very small* OR p_i is *very small*¹⁹) then the method is likely to fail.

¹⁹*very small*: the order of the local truncation error

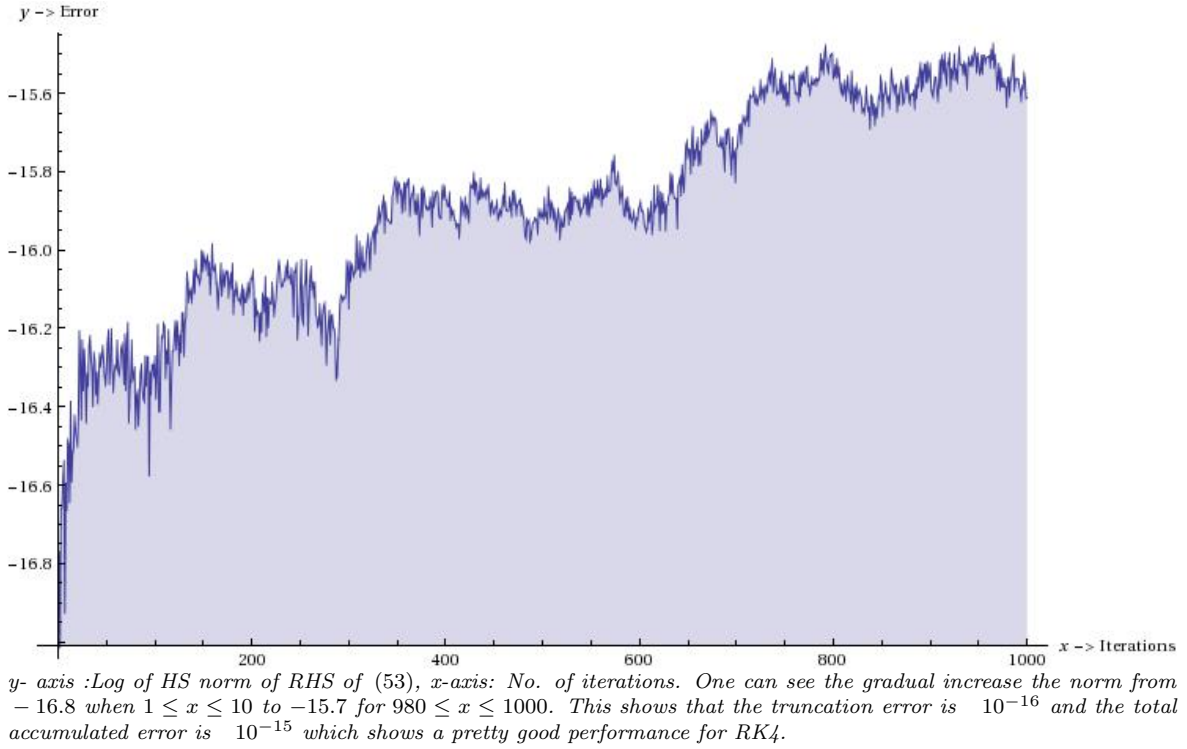


Figure 1: Error vs Iteration No.

5 Linearly Dependent States

This method generally won't work when for linearly dependent ensembles. Three major problems are identified:

IA. Why any such technique won't work generally: Optimal POVM doesn't always vary smoothly

From [21][22] we see that for an ensemble of linearly dependent pure states $\{\tilde{p}_i \geq 0, |\psi_i\rangle\langle\psi_i|\}_{i=1}^m$, the ensemble for which the PGM of the former will discriminate optimally is given by $\{p_i > 0, |\psi_i\rangle\langle\psi_i|\}_{i=1}^m$ satisfying the rule:

$$p_i \leq \frac{c}{\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle} \text{ when } \tilde{p}_i = 0$$

$$p_i = \frac{c}{\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle} \text{ when } \tilde{p}_i > 0$$

where c is some normalization constant and $\sqrt{\tilde{\rho}^{-1}}$ is the inverse of the positive square root for the ensemble $\{\tilde{p}_i, \psi_i\}$.

Our problem is opposite: we need to find the optimal POVM given the ensemble. But from the above theorem one can conclude that as one varies the ensemble of states, the optimal POVM will not necessary vary in a smooth fashion. Also, the above mapping isn't one-to-one; i.e. for the same ensemble one can find two or more ensembles whose PGM's will optimally discriminate among the members of the former ensemble.

IB. Why any such technique won't work generally: Optimal POVM can vary discontinuously

For linearly dependent pure states, the optimal POVM for an ensemble need not be unique. For example the optimal POVM for the equiprobable ensemble $\{\frac{1}{4}|0\rangle\langle 0|, \frac{1}{4}|1\rangle\langle 1|, \frac{1}{4}|+\rangle\langle +|, \frac{1}{4}|-\rangle\langle -|\}$ ²⁰ has two extremal optimal solutions - $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ and $\{|+\rangle\langle +|, |-\rangle\langle -|\}$ and hence any convex combination of these two is also an optimal solution. Were this ensemble to lie on a trajectory, the optimal solution of points on one side of this ensemble could be closer to $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ than $\{|+\rangle\langle +|, |-\rangle\langle -|\}$ and vice versa for points on the other side of this ensemble. This shows that the optimal POVM can vary discontinuously.

II. Why this technique won't work: \nexists a set of states $\{|u_i\rangle \mid \langle \tilde{\psi}_i | \tilde{u}_j \rangle = \delta_{i,j}\}$ when $\{|\psi_i\rangle\}_i$ are linearly dependent

Our technique relies upon casting the problem in a representation that exists only when the states are linearly independent. The speciality with this representation is that the one can use the implicit function theorem which demands that the number of function y_i and the number of implicit variables f_i have to be equal.

References

- [1] Helstrom, Carl W., Quantum Detection and Estimation
- [2] Barnett Stephen M, Phys. Rev. A. 64 030303(R), "Minimum Error Discrimination between Multiply Symmetric States"
- [3] Ban M., Kurokawa K., Momosa R., Hirota O., "Optimum Measurements for Discrimination Among Symmetric Quantum States and Parameter Estimation " Int. J. Theory. Phys 55 22(1987)
- [4] Chou C. L., Hsu L. Y., Phys. Rev. A 68, 042305, "Minimum Error Discrimination between Symmetric Mixed States"
- [5] Y. C. Eldar, A Megretski, G. C. Verghese, IEEE, Trans. Inf. Theory 50,1198(2004)
- [6] Kennedy, R. S., M. I. T. Res. Lab.: Electron. Quart. Progr. Rep. 110, 142 (1973)
- [7] Yuen H.P., Kennedy R. S., Lax M., IEEE Trans. Inform. Theory, IT-21, 125 (1975)
- [8] Holevo A. S., J. Multivariate Anal. 3, 337 (1973)
- [9] Ha Dhongon, Kwon Younghun Phys. Rev. A 87 062302 (2013) "Complete Analysis of Three Qubit Mixed States"
- [10] Bae J., New. J. Phys 15073037 (2013)
- [11] Samsonov Boris F., Phys. Rev. A 87 012334 (2009), "Minimum Error Discrimination of Qubit States"
- [12] Peres A, Terno D. R. "Optimal Distinction between non-orthogonal quantum states" J. Phys A 31, 7105-7111(1998)
- [13] Chefles A., Phys.Lett. A239 (1998) 339-347, "Unambiguous Discrimination Between Linearly-Independent Quantum States"
- [14] Pang S., Wu S., Phys. Rev. A, 80 052320(2009), "Optimum Unambiguous Discrimination of Linearly Independent Pure States"
- [15] Bergou J. A., Herzog U., Phys. Rev. A 71, 050301 (2005), "Optimum Unambiguous Discrimination of 2 Mixed Quantum States"
- [16] Raynal P., Lutkenhaus N., Phys. Rev A. 68, 022308 (2003), "Optimum Unambiguous Discrimination of Two Density Matrices: A Second Class of Exact Solutions"

²⁰ $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$

- [17] Herzog U., Phys. Rev. A 75 052309 (2007), "Optimum Unambiguous Discrimination of Two Mixed States and Application to a Class of Similar States"
- [18] Bergou J., Futschik F., Feldman E., Phys. Rev. Lett. 250502 (2012), "Optimum Unambiguous Discrimination of Pure Quantum States"
- [19] Croke S., Andersson E., Barnett S. M., Gilson C., Jeffers J., Phys. Rev. Lett. 96, 070401, (2006), "Maximum Confidence Measurements"
- [20] Goyal S., Simon N., Singh. R., Simon S., <http://arxiv.org/abs/1111.4427v1> "Geometry of the Generalized Bloch sphere for qutrit"
- [21] Mocho C., Phys. Rev. A 73, 032328, (2006), "Family of generalized "pretty good" measurements and the minimal-error pure-state discrimination problems for which they are optimal"
- [22] Belavkin V. P. Stochastics, 1, 315 (1975)
- [23] Jezek M., Rehacek J., Fiurasek J., Phys. Rev. A 65, 060301(R), 2002
- [24] Helstrom C. W., IEEE Trans. Inf. Theory IT-28, 359 (1982)
- [25] Chen C. , PhD. Thesis, <http://www.orcca.on.ca/~cchen/research.html>, "Solving Polynomial Systems via Triangular Decomposition "